

RAW SEQUENCE LISTING

EFS

The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.

Application Serial Number: 10/820,953A
Source: IFW/b
Date Processed by STIC: 4/5/07

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/820,953A

TIME: 15:53:41

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

```

3 <110> APPLICANT: Koelsch, Gerald
4      Tang, Jordan J. N.
5      Hong, Lin
6      Ghosh, Arun K.
7      The Board of Trustees of the University of Illinois
8      Oklahoma Medical Research Foundation
10 <120> TITLE OF INVENTION: Inhibitors of Memapsin 2 and Use Thereof
12 <130> FILE REFERENCE: 022266-000930US
C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/820,953A
C--> 15 <141> CURRENT FILING DATE: 2004-04-08
17 <150> PRIOR APPLICATION NUMBER: US 60/141,363
18 <151> PRIOR FILING DATE: 1999-06-28
20 <150> PRIOR APPLICATION NUMBER: US 60/168,060
21 <151> PRIOR FILING DATE: 1999-11-30
23 <150> PRIOR APPLICATION NUMBER: US 60/177,836
24 <151> PRIOR FILING DATE: 2000-01-25
26 <150> PRIOR APPLICATION NUMBER: US 60/178,368
27 <151> PRIOR FILING DATE: 2000-01-27
29 <150> PRIOR APPLICATION NUMBER: US 60/210,292
30 <151> PRIOR FILING DATE: 2000-06-08
32 <150> PRIOR APPLICATION NUMBER: US 09/603,713
33 <151> PRIOR FILING DATE: 2000-06-27
35 <150> PRIOR APPLICATION NUMBER: US 09/845,226
36 <151> PRIOR FILING DATE: 2001-04-30
38 <160> NUMBER OF SEQ ID NOS: 39
40 <170> SOFTWARE: PatentIn Ver. 2.1
42 <210> SEQ ID NO: 1
43 <211> LENGTH: 3252
44 <212> TYPE: DNA
45 <213> ORGANISM: Homo sapiens
47 <220> FEATURE:
48 <223> OTHER INFORMATION: memapsin 2
50 <400> SEQUENCE: 1
51 gcgggagtg cgcctgcccc cggcaccag cagggcatcc ggctgcccct gcgcagcggc 60
52 ctggggggcg cccccctggg gctgcggctg ccccgaggaga ccgacgaaga gcccgaggag 120
53 cccggccgga ggggcagctt tgtggagatg gtggacaacc tgagggggcaa gtcggggcag 180
54 ggctactacg tggagatgac cgtgggcagc cccccgcaga cgctcaacat cctgggtggat 240
55 acaggcagca gtaactttgc agtgggtgct gccccccacc ccttcctgca tcgctactac 300
56 cagaggcagc tgtccagcac ataccgggac ctccggaagg gtgtgtatgt gccctacacc 360
57 cagggcaagt gggaagggga gctgggcacc gacctggtaa gcatcccca tggccccaac 420
58 gtcactgtgc gtgccaacat tgctgccatc actgaatcag acaagttctt catcaacggc 480
59 tccaactggg aaggcatcct ggggctggcc tatgctgaga ttgccaggcc tgacgactcc 540
60 ctggagcctt tctttgactc tctggtaagg cagaccacg ttcccaacct cttctccctg 600

```

RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/820,953A

TIME: 15:53:41

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

```

61 cagctttgtg gtgctggcct cccctcaac cagtctgaag tgctggcctc tgtcggagg 660
62 agcatgatca ttggaggatg cgaccactcg ctgtacacag gcagtctctg gtatacacc 720
63 atccggcggg agtgggtatta tgagggtgatc attgtgcggg tggagatcaa tggacaggat 780
64 ctgaaaatgg actgcaagga gtacaactat gacaagagca ttgtggacag tggcaccacc 840
65 aaccttcgtt tgcccaagaa agtgtttgaa gctgcagtca aatccatcaa ggcagcctcc 900
66 tccacggaga agttccctga tggtttctgg ctaggagagc agctgggtgtg ctggcaagca 960
67 ggcaccaccc cttggaacat tttcccagtc atctcactct acctaatggg tgaggttacc 1020
68 aaccagtcct tccgcatcac catccttccg cagcaatacc tgcggccagt ggaagatgtg 1080
69 gccacgtccc aagacgactg ttacaagttt gccatctcac agtcatccac gggcactgtt 1140
70 atgggagctg ttatcatgga gggcttctac gttgtctttg atcggggccg aaaacgaatt 1200
71 ggctttgtctg tcagcgcttg ccagtgtcac gatgagttca ggacggcagc ggtggaaggc 1260
72 cttttgtca cttggacat ggaagactgt ggctacaaca tccacagac agatgagtca 1320
73 accctcatga ccatagccta tgtcatggct gccatctgcg cctcttcat gctgccactc 1380
74 tgctcatgg tgtgtcagtg gcgctgcctc cgctgcctgc gccagcagca tgatgacttt 1440
75 gctgatgaca tctcctgtct gaagtggaga ggccccatgg cagaagatag agattccctc 1500
76 ggaccacacc tccgtggttc actttggcca caagtaggag acacagatgg cacctgtggc 1560
77 cagagcacct caggaccctc cccaccacc aaatgcctct gccttgatgg agaaggaaaa 1620
78 ggctggcaag gtgggttcca gggactgtac ctgtaggaaa cagaaaagag aagaaagaag 1680
79 cactctgctg gcgggaatac tcttggtcac ctcaaattta agtcgggaaa ttctgctgct 1740
80 tgaacttca gccctgaacc tttgtccacc attcctttaa attctccaac ccaagtatt 1800
81 cttcttttct tagtttcaga agtactggca tcacacgcag gttaccttgg cgtgtgtccc 1860
82 tgtggtaccc tggcagagaa gagaccaagc ttgtttccct gctggccaaa gtcagtagga 1920
83 gaggatgcac agtttgctat ttgctttaga gacagggact gtataaacia gcctaacatt 1980
84 ggtgcaaaga ttgcctcttg aattaaaaaa aaactagatt gactatttat acaaatgggg 2040
85 gcggctggaa agaggagaag gagagggagt acaaagacag ggaatagtgg gatcaaagct 2100
86 aggaaaggca gaaacacaac cactcaccag tcctagtttt agacctcatc tccaagatag 2160
87 catcccatct cagaagatgg gtgtgtgttt caatgttttc ttttctgtgg ttgcagcctg 2220
88 accaaaagtg agatgggaag ggcttatcta gccaaagagc tcttttttag ctctcttaaa 2280
89 tgaagtgcc actaagaagt tccacttaac acatgaattt ctgccatatt aatttcattg 2340
90 tctctatctg aaccaccctt tattctacat atgataggca gcactgaaat atcctaacc 2400
91 cctaagctcc aggtgcctg tgggagagca actggactat agcagggtg ggctctgtct 2460
92 tctggctcat aggtcactc tttccccc aaatctctct ggagctttgc agccaagggtg 2520
93 ctaaaaggaa taggtaggag acctcttcta tctaactctt aaaagcataa tgttgaacat 2580
94 tcattcaaca gctgatgcc tataaccct gcctggattt ctctctatta ggctataaga 2640
95 agtagcaaga tctttacata attcagagt gtttcattgc ctctctacct tctctaattg 2700
96 cccctccatt tatttgacta aagcatcrca cagtggcact agcattatac caagagtatg 2760
97 agaaatacag tgctttatgg ctctaacatt actgccttca gtatcaaggc tgctggaga 2820
98 aaggatggca gcctcagggc ttccttatgt cctccaccac aagagctcct tgatgaagg 2880
99 catcttttcc cctatcctg tcttccct cccgctcct aatggtacgt gggtagccag 2940
100 gctgggtctt gggctaggta gtggggacca agttcattac ctccctatca gttctagcat 3000
101 agtaaactac ggtaccagtg ttagtgggaa gagctgggtt ttcctagtat acccactgca 3060
102 tctactcct acctggtcaa cccgtgctt ccaggtatgg gacctgctaa gtgtggaatt 3120
103 acctgataag ggagaggga atacaaggag ggcctctggg gttcctggcc tcagccagct 3180
104 gccmcaagc cataaacaa taaamcaaga atactgagtc taaaaaaaaa aaaaaaaaaa 3240
105 aaaaaaaaaa aa 3252
108 <210> SEQ ID NO: 2
109 <211> LENGTH: 488
110 <212> TYPE: PRT
111 <213> ORGANISM: Homo sapiens

```

RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/820,953A

TIME: 15:53:41

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

```

113 <220> FEATURE:
114 <223> OTHER INFORMATION: purified memapsin 2, aspartic proteinase 2 (ASP2)
116 <220> FEATURE:
117 <223> OTHER INFORMATION: amino acids 28-48 are remnant putative propeptide
118     residues
122 <220> FEATURE:
123 <223> OTHER INFORMATION: amino acids 54-57, 61-68, 73-80, 86-89, 109-111,
124     113-118, 123-134, 143-154, 165-168, 198-202, and
125     220-224 are N-lobe beta strands
127 <220> FEATURE:
128 <223> OTHER INFORMATION: amino acids 58-61, 78, 80, 82-83, 116, 118-121,
129     156, 166, 174, 246, 274, 276, 278-281, 283, and
130     376-377 are residues in contact with the OM99-2
131     inhibitor
133 <220> FEATURE:
134 <223> OTHER INFORMATION: amino acids 184-191 and 210-217 are N-lobe helices
136 <220> FEATURE:
137 <223> OTHER INFORMATION: amino acids 237-240, 247-249, 251-256, 259-260,
138     273-275, 282-285, 316-318, 331-336, 342-348,
139     354-357, 366-370, 372-375, 380-383, 390-395,
140     400-405, and 418-420 are C-lobe beta strands
142 <220> FEATURE:
143 <223> OTHER INFORMATION: amino acids 286-299, 307-310, 350-353, 384-387,
144     and 427-431 are C-lobe helices
146 <400> SEQUENCE: 2
147 Ala Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro
148   1             5             10             15
150 Leu Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg
151             20             25             30
153 Glu Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val
154             35             40             45
156 Glu Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val
157             50             55             60
159 Glu Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp
160   65             70             75             80
162 Thr Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu
163             85             90             95
165 His Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg
166             100            105            110
168 Lys Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu
169             115            120            125
171 Gly Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg
172             130            135            140
174 Ala Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly
175   145            150            155            160
177 Ser Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg
178             165            170            175
180 Pro Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr
181             180            185            190

```

RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/820,953A

TIME: 15:53:41

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

```

184 His Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro
185      195      200      205
187 Leu Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile
188      210      215      220
190 Gly Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro
191 225      230      235      240
193 Ile Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile
194      245      250      255
196 Asn Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys
197      260      265      270
199 Ser Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val
200      275      280      285
202 Phe Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys
203      290      295      300
205 Phe Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala
206 305      310      315      320
208 Gly Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met
209      325      330      335
211 Gly Glu Val Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln
212      340      345      350
214 Tyr Leu Arg Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr
215      355      360      365
217 Lys Phe Ala Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val
218      370      375      380
220 Ile Met Glu Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile
221 385      390      395      400
223 Gly Phe Ala Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala
224      405      410      415
226 Ala Val Glu Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr
227      420      425      430
229 Asn Ile Pro Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val
230      435      440      445
232 Met Ala Ala Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val
233      450      455      460
235 Cys Gln Trp Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe
236 465      470      475      480
238 Ala Asp Asp Ile Ser Leu Leu Lys
239      485
243 <210> SEQ ID NO: 3
244 <211> LENGTH: 503
245 <212> TYPE: PRT
246 <213> ORGANISM: Homo sapiens
248 <220> FEATURE:
249 <223> OTHER INFORMATION: pro-memapsin 2
251 <220> FEATURE:
252 <223> OTHER INFORMATION: amino acids 1-13 are the T7 promoter
254 <220> FEATURE:
255 <223> OTHER INFORMATION: amino acids 1-15 are vector-derived residues
257 <220> FEATURE:

```

DATE: 04/05/2007

TIME: 15:53:41

Input Set : N:\EFS\04 05 07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

```

258 <223> OTHER INFORMATION: amino acids 16-64 are a putative pro-peptide
260 <220> FEATURE:
261 <223> OTHER INFORMATION: amino acids 16-456 are pro-memapsin 2 T1
263 <220> FEATURE:
264 <223> OTHER INFORMATION: amino acids 16-421 are promemapsin 2 T2
266 <400> SEQUENCE: 3
267 Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Gly Ser Met Ala
268 1 5 10 15
270 Gly Val Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro Leu
271 20 25 30
273 Arg Ser Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg Glu
274 35 40 45
276 Thr Asp Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val Glu
277 50 55 60
279 Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val Glu
280 65 70 75 80
282 Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp Thr
283 85 90 95
285 Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu His
286 100 105 110
288 Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg Lys
289 115 120 125
291 Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly
292 130 135 140
294 Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg Ala
295 145 150 155 160
297 Asn Ile Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser
298 165 170 175
300 Asn Trp Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro
301 180 185 190
304 Asp Asp Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His
305 195 200 205
307 Val Pro Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu
308 210 215 220
310 Asn Gln Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly
311 225 230 235 240
313 Gly Ile Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile
314 245 250 255
316 Arg Arg Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn
317 260 265 270
319 Gly Gln Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser
320 275 280 285
322 Ile Val Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe
323 290 295 300
325 Glu Ala Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe
326 305 310 315 320
328 Pro Asp Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly
329 325 330 335
331 Thr Thr Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 04/05/2007
PATENT APPLICATION: US/10/820,953A TIME: 15:53:42

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt
Output Set: N:\CRF4\04052007\J820953A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:8; Xaa Pos. 2
Seq#:22; Xaa Pos. 3
Seq#:23; Xaa Pos. 1
Seq#:27; Xaa Pos. 3,4
Seq#:29; Xaa Pos. 2,3
Seq#:35; Xaa Pos. 4,5

VERIFICATION SUMMARY

DATE: 04/05/2007

PATENT APPLICATION: US/10/820,953A

TIME: 15:53:42

Input Set : N:\EFS\04_05_07\10820953a_efs\sequence.txt

Output Set: N:\CRF4\04052007\J820953A.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application Number
L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:442 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8 after pos.:0
L:633 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22 after pos.:0
L:652 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23 after pos.:0
L:717 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27 after pos.:0
L:758 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:29 after pos.:0
L:789 M:259 W: Allowed number of lines exceeded, <223> Other Information:
L:794 M:283 W: Missing Blank Line separator, <220> field identifier
L:916 M:283 W: Missing Blank Line separator, <220> field identifier
L:929 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:35 after pos.:0